SUNITED STATES DISTRICT COURT
WESTERN DISTRICT OF TEXAS
AUSTIN DIVISION

FRESHUB, INC., a Delaware Corporation, and FRESHUB, LTD., an Israeli Limited Liability Company, Plaintiffs, v. AMAZON.COM, INC., a Delaware Corporation, AMAZON DIGITAL SERVICES, LLC, a Delaware Limited Liability Company, PRIME NOW, LLC, a Delaware Limited Liability Company, and WHOLE FOODS MARKET SERVICES, INC., a Delaware Corporation, Defendants.	Case No. 1:19-cv-00885-ADA

OPENING CLAIM CONSTRUCTION BRIEF OF AMAZON.COM, INC., AMAZON DIGITAL SERVICES, LLC, PRIME NOW, LLC, AND WHOLE FOODS MARKET SERVICES, INC.

J. David Hadden (CSB No. 176148)
Email: dhadden@fenwick.com
Saina S. Shamilov (CSB No. 215636)
Email: sshamilov@fenwick.com
Ravi R. Ranganath (CSB No. 272981)
Email: rranganath@fenwick.com
Vigen Salmastlian (CSB No. 276846)
Email: vsalmastlian@fenwick.com
FENWICK & WEST LLP
Silicon Valley Center
801 California Street
Mountain View, CA 94041
Telephone: 650.988.8500
Facsimile: 650.938.5200

Of Counsel:

Barry K. Shelton (TX Bar #24055029)
bshelton@sheltoncoburn.com
SHELTON COBURN LLP
311 RR 620 S, Suite 205
Austin, TX 78734
Tel: (512) 263-2165
Fax: (512) 263-2166

Counsel for Defendants
AMAZON.COM, INC.,
AMAZON DIGITAL SERVICES, LLC,
PRIME NOW, LLC, and
WHOLE FOODS MARKET SERVICES, INC.

January 24, 2020

TABLE OF CONTENTS

Page

TABLE OF AUTHORITIES

Page(s)

CASES

STATUTES

OTHER AUTHORITIES

I. INTRODUCTION

Freshub's patents disclose a refrigerator. With that refrigerator and a few aspirational paragraphs in a specification left abandoned for years, Freshub deigns to usurp six decades of ongoing academic and industry research chasing one of the most elusive results in artificial intelligence—"recognition and understanding of spontaneous and unrehearsed speech,"[1] *i.e.,* spoken language processing.

Seventy years ago, Alan Turing famously proposed that the ability for computers to process spoken language proficiently was the test of true artificial intelligence.[2] Since then, "the goal of building commercially viable spoken language systems has long attracted the attention of scientists and engineers all over the world."[3] In the last decade, many scientists and engineers, including those at Amazon, have made tremendous progress toward this elusive goal with cutting-edge deep learning AI algorithms.[4] Those algorithms enabled Amazon's Alexa, and Amazon scientists have published hundreds of scientific papers detailing their breakthroughs.[5] None of these breakthroughs came from Freshub's abandoned refrigerator patent.

Indeed, it was only *after* Amazon launched Alexa that Freshub revived its abandoned application, filed new claims dropping the refrigerator and claiming instead the idea of voice shopping with an unspecified voice processing system, rushed prosecution through the Patent Office, and filed this lawsuit claiming that *it* invented Alexa. But it did not. And its patents confirm so; they are devoid of any spoken language processing solution. Instead, they recite and claim a mere

---

[1] Declaration of J. David Hadden ("Hadden Decl."), Ex. 6 (Xuedong Huang, et al., "Spoken Language Processing: A Guide to Theory, Algorithm, and System Development," (Prentice Hall PTR, 2001)) ("SLP") at xxi–xxiii.

[2] *Id.* at 3; Hadden Decl., Ex. 8.

[3] SLP at 2.

[4] *See, e.g.*, Hadden Decl. Ex. 9.

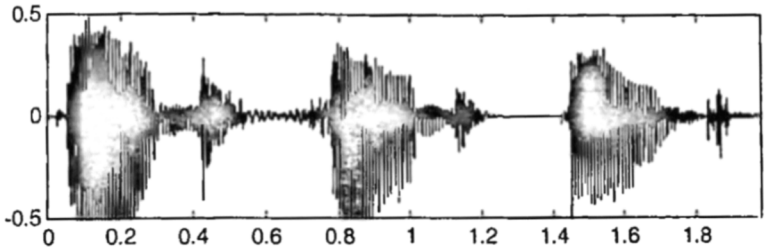
[5] *See id.* Ex. 10.

wish: translate human speech, understand it, and do what a human requested.

The integrity of our patent system is based on the fundamental bargain that a patent is awarded only in exchange for contributing to the public a specific new solution to a specific problem. One can only claim what one invented and "no more." *MySpace, Inc. v. GraphOn Corp.*, 672 F.3d 1250, 1256 (Fed. Cir. 2012). Freshub, however, "has in effect claimed everything that [processes spoken language] under the sun," thereby usurping a decade of work by thousands of Amazon researchers, six decades of computer science history that came before, and countless breakthroughs and innovations to come, all while contributing none to the public knowledge. Such claims are indefinite. *ePlus, Inc. v. Lawson Software, Inc.*, 700 F.3d 509, 519 (Fed. Cir. 2012).

II. OVERVIEW OF SPOKEN LANGUAGE PROCESSING

For a computer to respond to spoken human language is nearly magical: the input is a sound wave, a squiggly line on a computer screen.[6]



The output is a human intent that can be fulfilled—a request to play a song, to purchase a stock, to send an email, to answer a question. That output is not encoded in the sound wave; one cannot simply extract it like a sound from a radio wave or an image from a television signal. That output is determined by a complex system for processing spoken language, which requires speech recognition and natural language processing.[7] Both are difficult and the subject of intense ongoing research and development.

[6] SLP at 51(excerpt of Fig. 2.24).
[7] *Id*. at 2.

In the first process, dubbed Automatic Speech Recognition (ASR), a system converts the received sound wave to text.[8] To accomplish this, the sound wave is first digitally processed to extract the features most relevant for distinguishing sounds in human speech.[9] This requires an understanding of both the human vocal system and human auditory perception.[10] Next, the process attempts to map the extracted features to the most basic word constituents like phonemes.[11] Though there are millions of English words, they are all made up of fifty basic phonemes.[12] Researchers figured out that mapping to phonemes or other word constituents, which are also present in other languages, allows for a more efficient speech recognition process.[13] But using basic units such as phonemes creates other problems to solve.[14] Due to the structure and imperfections of the human vocal system, the actual sound of a spoken phoneme depends significantly on the surrounding phonemes and whether the phoneme is at the beginning, middle, or end of a word.[15] To address this issue, system designers will often use collections of sequential phonemes like triphones (three sequential phonemes) as the basic units to recognize.[16] But because there are many more triphones than phonemes, this again creates complexities in designing and training the system.[17]

Thus, the mapping of the series of features extracted from the audio signal to the most probable series of triphones (or phonemes or some other word sub-unit) is generally performed

[8] *Id*. at 4.
[9] *Id*. at 5.
[10] *Id.* at 4–5.
[11] *Id.* at 19–23.
[12] *Id.* at 36–38.
[13] *Id.* at 4, 19–23, 51.
[14] *Id.* at 68–69.
[15] *Id.* at 430.
[16] *Id.* at xxii, 430–36, 658, 808.
[17] *Id.*

using an acoustic model and a complex statistical classification algorithm. "Acoustic models include the representation of knowledge about acoustics, phonetics, microphone and environment variability, gender and dialect differences among speakers, etc."[18] At the acoustic phonetic level, this variability is typically modeled using statistical techniques applied to large amounts of data and by algorithms that adapt speaker-independent models to those of the current speaker during use.[19] Statistical data-driven classification algorithms use statistics or probabilities to classify data into categories.[20] Researchers have and continue to develop sophisticated classification algorithms including Hidden Markov Models, Neural Networks, and Finite State Transducers among others that are trained using machine learning algorithms.[21]

The output of the initial classifier algorithm, a series of phonemes or other word constituents, is then mapped to the most probable sequence of corresponding words using a language model and another machine learned statistical classifier. "Language models refer to a system's knowledge of what constitutes a possible word, what words are likely to co-occur, and in what sequence."[22] All of this is to map voice to text. That is the "easy" part.

Natural language understanding ("NLU"), which requires figuring out what the user intended by that sequence of words and what the computer can do to fulfill that intent, is the "hard" part. This is an immensely difficult problem that remains far from solved. "In comparison to

---

[18] *Id*. at 4.

[19] Hadden Decl. Ex. 7 (Ronald Cole, et al. (1997), *Survey of the State of the Art in Human Language Technology*, Cambridge University Press, ISBN 0-521-59277-1) ("Cole") at 6.

[20] SLP at 4; *see also* Cole at 342–348.

[21] Machine learning is the study of algorithms and statistical models that computers run to perform a specific task without using explicit instructions. Instead, they rely on patterns and inferences. As a scientific tool, machine learning turns information into knowledge. It can find valuable underlying patterns within complex training data sets that are otherwise a struggle to discover. *See* Hadden Decl. Ex. 11 at vii, xiii–xix, 12–32, 605–645; *see also* SLP at xxi–xxii 1–11, 19–53, 585, 763; Cole at xi–xii, 2, 5, 30, 73–74, 83, 116–118, 123, 170, 356, 368, 381.

[22] SLP at 4.

speech recognition and text to speech, spoken language understanding is further away from approaching the level of humans, especially for general purpose spoken language applications."[23]

At a high level, NLU, like ASR, is an inordinately complex pattern recognition problem. The goal is to map a sequence of words to an action to be performed by a computer.[24] But the patterns are far more elusive. "Whether a speaker is inquiring about flights to Seattle, reserving a table at a Pittsburgh restaurant, dictating an article in Chinese, or making a stock trade, a spoken language understanding system is needed to interpret utterances in context and to carry out appropriate actions."[25] But, to perform that mapping, the system must have a robust semantic model to know that Seattle is a city and a potential destination, that a table is an entity that can be reserved at a restaurant, that an article is a collection of words, and that a stock is a specific item that can be bought and sold at a current price. "Knowledge of the characteristic vocabulary, typical syntactic patterns, and possible actions in any given application context for both interpretation of user utterances and planning system activity are the heart and soul of any spoken language understanding system."[26]

These are the deep, complex problems computer and cognitive researchers have struggled with for over half a century. They are the same problems Freshub tries to bypass, stepping to the front of the line to claim the rewards without offering any solutions.

III. THE FRESHUB PATENTS

Freshub currently asserts 55 claims from four related patents—Nos. 9,908,153 (the "'153 patent"); 10,213,810 (the "'810 patent"); 10,232,408 (the "'408 patent"); and 10,239,094 (the

[23] *Id*. at 13.
[24] *Id*. at 854.
[25] *Id*. at 7.
[26] *Id*.

"'094 patent") (collectively, the "asserted patents"). The asserted patents share an identical specification and are all titled "Systems and methods for scanning information from storage area contents." (*See, e.g.,* '153 patent 1:1–14:45.). They all claim priority to U.S. Patent No. 9,821,344 (the "'344 patent"), which shares the same specification and title. (*See generally* '344 patent.)

The patents claim to address the problem of "faintly or poorly printed" expiration dates on products that are "often densely packed into a refrigerated storage unit" preventing users from "conveniently monitor[ing] th[ose] expiration dates." ('153 patent at 1:20–26.) To address this problem, the patents propose a system for reading tags on refrigerated items and prompting the user to use or replace the items before their expiration dates. (*Id*. at 1:35–37; Abstract.) Accordingly, the '344 patent claims an "electronic refrigeration system" that captures digital images of the stored refrigerated items and alerts users of any upcoming expiration dates. ('344 patent, cls. 1–7). Its four asserted children, however, claim a different subject matter. The prosecution history sheds light on why.

The '344 patent was filed in 2005. Its prosecution, however, was abandoned in 2012. Five years later—after Amazon released its revolutionary Alexa and Echo smart speakers—Freshub revived the application and rushed several continuation applications through the Patent Office on an expedited prosecution basis, which eventually issued as the patents in this case. (*Compare* '344 patent *to* '153, '408, '810, '094 patents (priority date Dec. 12, 2005).) Their claims do not claim a refrigerator. Instead, they are directed to the results of spoken language processing with no solutions on how to accomplish them. Claim 1 of the '153 patent is representative and recites:

> 1. A voice processing system comprising:
>
> a first system configured to receive user spoken words comprising:
>
> a microphone;
>
> a wireless network interface;

> a digitizer coupled to the microphone, wherein the digitizer is configured to convert spoken words into a digital representation;
>
> a first computer;
>
> non-transitory memory that stores instructions that when executed by the first computer cause the first system to perform operations comprising:
>
> receive via the digitizer a verbal order, comprising at least one item, from a user, wherein the verbal order was captured by the microphone and digitized by the digitizer;
>
> immediately transmit, using the wireless network interface, the digitized order to a computer system remote from the first system;
>
> the computer system, the computer system comprising:
>
> a networks interface;
>
> a second computer;
>
> non-transitory memory that stores instructions that when executed by the second computer cause the computer system to perform operations comprising:
>
> receive, using the network interface, the digitized order from the first system;
>
> translate at least a portion of the digitized order to text;
>
> identify an item corresponding to the text;
>
> add the identified item to a list associated with the user;
>
> enable the list, including the identified item, to be displayed via a user display.

('153 patent cl. 1.) In other words, claim 1 recites a system that: (1) receives a spoken order from a user; (2) translates the voice input to text; (3) identifies the item the user requested; and (4) adds the item to a user's set, or "list" for display. (*Id.*) The other independent claims in each of the asserted patents recite the same results. (*See also, e.g.*, '408 patent cls. 17, 30; '810 patent, cls. 1, 17; '094 patent, cls. 1, 20; '153 patent, cl. 1.) The only conceivable support for these claims in the

specification is less than a column of description that relies on Figure 8 and is repeated with slight variation at places in the specification. (*See, e.g.*, '153 patent, 8:17–9:14; 13:56–14:44.)

At the first step of Figure 8, the user activates a button or uses a voice command, *e.g.*, "put milk on my shopping list." ('153 patent at 8:9–11, 8:21–25.) The system uses a "voice recording device[]" with a digital or analog memory to record the user's request. (*Id.* at 8:9–11, 17–25, 35–36.) Next, the system transmits the user's recorded request to a "remote system" of Figure 2. (*Id.* at 14:10–12.) The next step—"perform voice recognition"—is a key result. The specification discloses no algorithm, method, flow chart, or other explanation of how the claimed system achieves it. (*Id.* at 14:15–17.) Nor do the claims recite any solution or improvement to the plethora of potential algorithms and computing models developed over decades of research on spoken language processing. Instead, Freshub collapses mind-bendingly complicated ASR and NLU processes into unspecified black boxes and claims a result that usurps the past and future hard work and innovation of others. Freshub's patents claim the idea of voice shopping, not a solution for achieving it. *See, e.g., O'Reilly v. Morse*, 56 U.S. (15 How.) 62, 112 (1853).

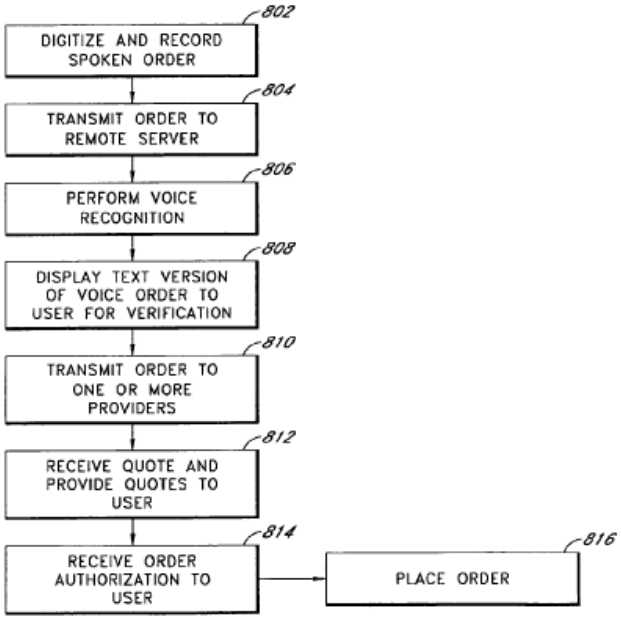


FIG. 8

IV. THE DISPUTED CLAIM TERMS AND PHRASES

Out of the 55 asserted claims, the parties dispute constructions of 16 claim terms. The disputed terms fall generally into two buckets: (1) terms that are subject to § 112 ¶ 6 but lack any corresponding structures, and (2) terms that are ambiguous even to one skilled in the art, due either

to their inconsistent usage across claims or because they lack antecedent basis.

A. Terms Governed by Section 112 ¶ 6

Our patent laws do not allow claims directed to a result. That rule was established in the 1800s and has been invoked repeatedly since. *See O'Reilly*, 56 U.S. (15 How.) at 120 ("no patent can lawfully issue" for a mere "effect . . . distinct from the process or machinery necessary to produce it"). Result-focused claims epitomize what the Federal Circuit has described as the "vice of functional claiming," where "the inventor is painstaking when he recites what has already been seen, and then uses conveniently functional language at the *exact point of novelty.*" *Halliburton Energy Servs., Inc. v. M-I LLC*, 514 F.3d 1244, 1255 (Fed. Cir. 2008) (quoting *Gen. Elec. Co. v. Wabash Appliance Corp.*, 304 U.S. 364, 371 (1938)) (emphasis added). That is precisely what Freshub did by concocting claims years *after* Alexa was invented and launched.

Fortunately, our patent laws stand in the way of such rent-seeking. Indeed, Congress enacted § 112 ¶ 6 to limit functional claiming like that employed by Freshub. *See Williamson v. Citrix Online, LLC*, 792 F.3d 1339, 1349–51 (Fed. Cir. 2015) (en banc) (imposing new standard of indefiniteness to address "proliferation of functional claiming untethered to § 112, para. 6 and free of the strictures set forth in the statute"); *Function Media, LLC v. Google, Inc.*, 708 F.3d 1310, 1319 (Fed. Cir. 2013). It allowed the patentee to claim the invention by the functions it performs, but limited the scope of those claims to the specific solutions disclosed in the specification—because in this country, the patentee cannot claim more than what she invented. *MySpace,* 672 F.3d at 1256.

In *Williamson*, the Federal Circuit reinforced and strengthened the longstanding prohibition against functional claiming: if the *claims* do not identify the specific structure or acts sufficient to achieve a claimed result—like processing spoken language—then the *specification* must fill the void to limit the claims to the inventor's specific solution. *Williamson*, 792 F.3d at 1350.

And this is true whether the claims recite the phrase "means for" or not. *Id.* at 1349–50. If the claims merely recite functions performed by unspecified black boxes, those claims are subject to the requirements of § 112 ¶ 6. *Id.* at 1350–51.

Many of the asserted claims in this case recite black boxes—*e.g.*, "computer," "computer system," "module," "processing device"—"configured to" perform or with "instructions that when executed" carry out the claimed functions. The phrases "configured to" and "instructions that when executed" make clear that *something else* has to be done *to* the black box for it to achieve the claimed functional result. The claims, however, are entirely silent as to how the claimed aspirations are to be achieved, as is the specification.

/ / /

/ / /

/ / /

/ / /

/ / /

/ / /

/ / /

/ / /

/ / /

/ / /

/ / /

/ / /

/ / /

/ / /

1. "Translate" functional terms are indefinite under section 112 ¶ 6.

Term	Amazon's Construction
"*instructions that when executed by the second computer cause the computer system to perform operations* comprising: . . . *translate* at least a portion of the digitized order to text" Claim 1 of the '153 patent	*Governed by pre-AIA 35 U.S.C. § 112 ¶ 6. Indefinite under pre-AIA 35 U.S.C. § 112* ***Functions:*** "translate at least a portion of the digitized order to text"
"*instructions that when executed by the computer cause the computer to perform operations* comprising: . . . *translate* at least a portion of the digitized [order / voice communication] to text" Claim 1 of the '801 patent; claims 1 and 30 of the '408 patent	"translating, using a processing system comprising at least one processing device and configured to perform translation of voice orders to text, at least a portion of the digitized order to text"
"*translating, using a processing system comprising at least one processing device and configured to perform translation* of voice orders to text, at least a portion of the digitized order to text" Claim 20 of the '408 patent	translating, using a translation module executed by the second computer system, at least a portion of the digitized spoken order to text"
"*translating, using a translation module* executed by the second computer system, at least a portion of the digitized spoken order to text" Claim 1 of the '094 patent	"translate at least a portion of the digitized spoken order to text"
"*instructions that when executed by the first computer cause the first system to perform* operations comprising: . . . *translate* at least a portion of the digitized spoken order to text" Claim 20 of the '094 patent	***Structures:*** None.

All of the asserted claims attempt to capture the result of ASR—"translating" voice to text. According to those claims, that result is achieved by generic computing devices—a "computer," "system," "device," or "module"—alone or by executing unspecified "instructions."

The terms "[computer/processing] system," "[processing] device," and "[translation] module" are not structures sufficient to perform the claimed translation function. They are mere placeholders for a "means" and are insufficient to avoid the application of § 112 ¶ 6. Indeed, calling a

module for translation a *translation* module is just defining it by the function it performs. *Williamson*, 792 F.3d at 1350 (holding term "distributed learning control module" was governed by § 112 ¶ 6 because "the word 'module' . . . sets forth the same black box recitation of structure for providing the same specified function as if the term 'means' had been used"); Manual of Patent Examining Procedure § 2181 at ¶ I.A (identifying "system for," "device for," and "module for" as "non-structural generic placeholders" that invoke § 112 ¶ 6). The Federal Circuit has repeatedly held that claims that recite such black boxes with nothing more are subject to § 112 ¶ 6. *See Williamson*, 792 F.3d at 1350 (citing "black box recitation of structure" insufficient to avoid application of § 112 ¶ 6); *Robert Bosch, LLC v. Snap-On Inc.*, 769 F.3d 1094, 1099 (Fed. Cir. 2014) (construing "program recognition device" and "program loading device" under § 112 ¶ 6 and holding claims indefinite); *Diebold Nixdorf, Inc. v. Int'l Trade Comm'n*, 899 F.3d 1291, 1296–98 (Fed. Cir. 2018) (construing functional black box term "cheque standby unit" under § 112 ¶ 6); *Advanced Ground Info. Sys., Inc. v. Life360, Inc.*, 830 F.3d 1341, 1347–48 (Fed. Cir. 2016) (construing "symbol generator" under § 112 ¶ 6); *Blackboard, Inc. v. Desire2Learn, Inc.*, 574 F.3d 1371, 1384–85 (Fed. Cir. 2009) (construing "access control manager" under § 112 ¶ 6). So have district courts around the country. *See Verint Sys. Inc. v. Red Box Recorders Ltd.*, 166 F. Supp. 3d 364, 381 (S.D.N.Y. 2016) ("The term 'system,' although qualified as a 'communication monitoring system,' does not impart a sufficient structure. 'System' standing alone is a nonce word that does not describe a structure that could perform the listed functions and the modifier . . . provides a functional description of the system but no structure."); *Intellectual Ventures II LLC v. FedEx Corp.*, No. 16-cv-980-JRG, 2017 WL 5896180, at *33 (E.D. Tex. Nov. 29, 2017) ("data modification tool" depicted in patent figures as a "black box contained within" did not recite sufficient structure to save claim term from indefiniteness); *St. Isidore Research, LLC v. Comerica Inc.*, No.

15-cv-1390-JRG-JSP, 2016 WL 4988246, at *14 (E.D. Tex. Sept. 19, 2016) (construing claim reciting "processor" under § 112 ¶ 6 because the claimed processor was "defined only by the function that it performs," and the claim did not "describe how the [claimed] processors interact with each other or with other limitations in the claim to achieve their objectives").

And adding generic "instructions" to the black boxes does not change the § 112 ¶ 6 analysis. Indeed, the claims lack any details about the "instructions" used to program the generic systems, modules, and devices to perform the critical function of "translating" voice to text—a problem with which top computer scientists have struggled for decades. The translation function the patents claim is an immensely complex operation requiring a plethora of design and algorithm choices, none of which is described in the claims, which recite only the desired and difficult result. Such claims fall under section 112 ¶ 6. *See Global Equity Mgmt. (SA) Pty. Ltd. v. Expedia, Inc.*, No. 2: 16-cv-00095-RWS-RSP, 2016 WL 7416132, at *27–29 (E.D. Tex. Dec. 22, 2016) ("*GEMSA*").

In *GEMSA*, the court considered the construction of the term "program code for configuring said at least one partition of said at least one secondary storage device through a secondary storage partitions window." *Id.* The court concluded that the term should be construed under § 112 ¶ 6 because the term was "defined only by the function that it performs," and the claims neither recited the "objectives and operations" of the code nor specified "[h]ow the code interacts with other code." *Id.* at *29. The same is even more true here. *See Cypress Lake Software, Inc. v. Samsung Elecs. Am., Inc.*, 382 F.Supp.3d 586, 615 (E.D. Tex. 2019) (claims reciting "code for" performing the functions of "detecting" user input and "presenting" a "navigation control" or a "visual component" governed by § 112 ¶ 6 because "the term 'code for' is defined only by the function that it performs"); *Cypress Lake Software, Inc. v. ZTE (USA)*, No. 17-cv-003000-RWS,

2018 WL 4035968, at *9 (E.D. Tex. Aug. 23, 2018) ("the term 'code for' does not connote sufficiently definite structure"). Freshub's claims provide no description as to how the "instructions" operate to achieve the elusive result or how the other black boxes "are configured" to do the same. Nor does the specification.[27]

The specification provides no specific protocols, no flow-charts, and no other description of how the instructions execute or implement the function of translating a "digitize[d] and record[ed] spoken order" (Fig. 8, State 802) into a "text version of voice order" (Fig. 8, State 808). Instead, it merely suggests that the translation function can be achieved using "grammar constrained recognition and/or natural language recognition." All this says is that you could try to solve the hard problem of building an ASR system by recognizing only limited combinations of words ("grammar constrained recognition")—some way, somehow, or the even harder problem of recognizing any set of words ("natural language recognition")—some way, somehow. Neither the way nor the how is described; the specification provides no solution for either. ('810 at 14:13–19.) This is akin to claiming a supersonic aircraft and stating that it can be achieved by merely flying faster than the speed of sound. Such a proclamation is an aspiration; it is not a solution. *See, e.g.*, *Aristocrat Techs.*, *Austl. PTY Ltd. v. Int'l Game Tech.*, 521 F.3d 1328, 1335 (Fed. Cir. 2008) (describing "the results of the operation of an unspecified algorithm" is insufficient to transform a general-purpose computer into structure sufficient to satisfy § 112 ¶ 6); *Media Rights Techs., Inc. v. Capital One Fin. Corp.*, 800 F.3d 1366, 1374 (Fed. Cir. 2015) (claim reciting "compliance mechanism" invalid under § 112 ¶ 6 where "the specification fails to disclose an operative

[27] Where neither the claims nor specification disclose structure, no expert testimony is necessary to conclude that a claim term is governed by § 112 ¶ 6. *See, e.g.*, *Noah Sys., Inc. v. Intuit Inc.*, 675 F.3d 1302, 1318–19 (Fed. Cir. 2012) (expert testimony cannot supplant absence of structure in the patent itself).

algorithm" for the claimed functions); *Twin Peaks Software Inc. v. IBM Corp.*, 690 F. App'x 656, 660–61 (Fed. Cir. 2017) (claim reciting "means for mounting" invalid under § 112 ¶ 6 where specification did not disclose an algorithm, and instead "circularly described" the "results of various 'mount' operations, rather than . . . disclosing the identity of the '*means* for mounting.'") (emphasis in original); *Grecia v. Samsung Elecs. Am., Inc.*, 780 F. App'x 912, 916–917 (Fed. Cir. 2019) (claim reciting "customization module" indefinite under § 112 ¶ 6 where "specification merely describes the results of customization without any algorithm for configuring the claimed module to obtain those results"); *Arendi S.A.R.L. v. LG Elecs., Inc.*, No. CV 12-1595-LPS, 2019 WL 3891150, at *11 (D. Del. Aug. 19, 2019) (means-plus-function terms indefinite where "specification fails to disclose any actual *algorithm*—whether in prose, as a mathematical formula, as a flow chart, or in any other suitable format—that could be followed to determine *which* *text*, of all of the text in a document, is a name or address.") (underlining emphasis added); *Cypress Lake Software, Inc.*, 2018 WL 4035968 at *10–11 ("There is no algorithm described in any form for the function of 'detecting the user input corresponding to the first navigation control.' Instead, the specification merely provides functional language and does not contain any process for detecting the user input."); *Auto-Dril, Inc. v. Nat'l Oilwell Varco, LP.*, 304 F. Supp. 3d 587, 618 (S.D. Tex. 2018), as amended (Apr. 20, 2018), *aff'd sub nom. Auto-Dril, Inc. v. Nat'l Oilwell Varco, L.P.*, 757 F. App'x 1012 (Fed. Cir. 2019) (claim reciting "electronic bit weight comparison means" and "computer data and program processing means" indefinite where "[t]he specification does not disclose any algorithm to perform the claim functions," such that "[o]ne skilled in the art of programming would have to create the function-performing algorithm"). Indeed, the functions claimed in the patents invalidated in each of these cases are trivial compared to the ASR result that Freshub claims.

2. "Matching" and "identifying" functional terms are indefinite under section 112 ¶ 6.

The asserted claims also recite the functions of matching the translated text to a description in a database and identifying an item corresponding to the translated texts. The claims again provide no solution for achieving this result. Nor does the specification.

Yet the claimed functions are difficult to implement and require sophisticated NLU processes. Once a machine has converted sound waves to text, it must still determine what the user intended by the spoken words or phrases to be of any use. This NLU process is even more difficult than determining words from spoken sounds and the known "solutions" remain in flux and vary greatly. "In comparison to speech recognition and text to speech, spoken language understanding is further away from approaching the level of humans, especially for general purpose spoken language applications." (SLP at 13.) For example, a person says: "I want a frozen peach pie." To get the order correct, the system would have to "understand" *at least* that (1) the phrase "I want" indicates that the user intends to make a purchase; (2) the item the user wants to purchase is a *pie*, and not a movie called "*frozen*" or a *frozen peach* or *peaches*; (3) *peach* is an available flavor or type of *pie*; (4) *frozen* is an attribute that modifies *pie*; and (5) the modifier *frozen* places limits on how the pie can be picked up or delivered. Neither the claims nor the specification in Freshub's patents disclose an algorithm or other structure that is capable of achieving any of this. The matching and identifying functional claim terms are governed by § 112 ¶ 6 and are indefinite.

a. The "matching" functional terms

Term	Amazon's Construction
"*instructions that when executed by the computer cause the computer to perform operations* comprising: . . . *match* the text, translated from the digitized order, to a text description stored in a database comprising text descriptions of items and associated unique product identifiers"	*Governed by pre-AIA 35 U.S.C. § 112 ¶ 6. Indefinite under pre-AIA 35 U.S.C. § 112* ***Functions:***

Claim 1 of the '408 patent "*matching*, *using the processing system*, the text, translated from the digitized order, to a text description associated with a unique product identifier" Claim 20 of the '408 patent "*instructions that when executed by a computer cause the computer to perform operations* comprising: . . . *match* the text, translated from the digitized voice communication, to a text description associated with a unique product identifier, wherein the text description is accessed from a data store" Claim 30 of the '408 patent "*instructions that when executed by the first computer cause the first system to perform operations* comprising: . . . *match* the text, translated from the digitized spoken order, to text descriptions of items, wherein the text descriptions of items are stored in a data store" Claim 20 of the '094 patent	"match the text, translate from the from the digitized order, to a text description stored in a database comprising text descriptions of items and associated unique product identifiers" "matching, using the processing system, the text, translated from the digitized order, to a text description associated with a unique product identifier" "match the text, translated from the digitized voice communication, to a text description associated with a unique product identifier, wherein the text description is accessed from a data store" "match the text, translated from the digitized spoken order, to text descriptions of items, wherein the text descriptions of items are stored in a data store" ***Structures:*** None

Similar to the translation terms, the claims again recite either a generic "processing system" or a generic "computer" executing unspecified instructions as performing the claimed matching function. And, again, as described above, neither is a sufficient structure for performing the claimed NLU matching function—a result that has been and continues to be the focus of intense research by computer scientists, linguists, and engineers. (*See* Sec. IV.B.1); *Williamson*, 792 F.3d at 1350 (functional black boxes do not disclose sufficient structure); *Verint Sys. Inc.,* 166 F. Supp. 3d at 381 ("system" is not sufficient structure to avoid application of § 112 ¶ 6); *St. Isidore*, 2016 WL 4988246, at *14 (generic "processor" insufficient structure because it is "defined only by the function that it performs"); *GEMSA*, 2016 WL 7416132 at *27–29 (claim reciting "program code"

for performing function governed by § 112 ¶ 6 and indefinite); *Cypress Lake Software, Inc.*, 382 F. Supp. 3d at 615 (claim reciting "code for" performing function governed by § 112 ¶ 6 and indefinite).

The specification provides no solution for matching the translated text to the items in the database either. It merely repeats the function itself. But to match the translated text to the items in the database is not a trivial process. For example, it requires an algorithm that, when it receives a request for a "*frozen peach pie,*" can understand that "*peach frozen pie,*" "*pie—frozen peach,*" "*freezer ready peach flavored pie,*" or "*frozen nectarine pie*" are potential "matches" for the user's spoken command, and mis-hits like the movie "*Frozen*," "*package of sliced frozen peaches*," and "*peach pie recipes*" are not. Instead of providing a solution to the problem, Freshub merely repeats the claimed function itself. (*See, e.g.*, '408 patent at 8:43–47.) And Figure 8 skips this key step entirely. (*See id.*, Fig. 8.) Instead of disclosing *how* to match the translated text with database product identifiers, the specification does exactly what the Supreme Court prevented Samuel Morse from doing—it attempts to capture any conceivable means or methodology for achieving the "matching" result. These § 112 ¶ 6 terms are indefinite.

b. The "identifying" functional terms

Term	Amazon's Construction
"instructions that when executed by the second computer cause the computer system to perform operations comprising: . . . *identify* an item corresponding to the text" Claim 1 of the '153 patent "instructions that when executed by the computer cause the computer to perform operations comprising: . . . *identify* an item corresponding to the text [description]" Claim 1 of the '810 patent; Claims 1 and 30 of the '408 patent	*Governed by pre-AIA 35 U.S.C. § 112 ¶ 6. Indefinite under pre-AIA 35 U.S.C. § 112* ***Functions:*** *"*identify an item corresponding to the text" "identify an item corresponding to the text [description]"

"instructions that when executed by the first computer cause the first system to perform operations comprising: . . . based on at least an identified match, *identify* a corresponding item" Claim 20 of the '094 patent	"identify a corresponding item" ***Structures:*** None

Some of the asserted claims recite the NLU result using even more generic language, *e.g.*, "identify an item corresponding to the text." As the other functional terms, the claimed "identify[ing]" is performed either by a generic processing "system" or "computer" executing unspecified instructions. Neither the system nor the instructions are structures sufficient to perform the function of identifying an item corresponding to the translated text. (*See* Sec. IV.B.1); *Williamson*, 792 F.3d at 1350; *Verint Sys. Inc.,* 166 F. Supp. 3d at 381 (construing "system" under § 112 ¶ 6); *GEMSA*, 2016 WL 7416132 at *27-29 (construing "program code" under § 112 ¶ 6); *Cypress Lake Software, Inc.*, 382 F. Supp. 3d at 615 (construing "code for" under ¶ 112 ¶ 6).

Indeed, the identifying function is merely a more generic variant of the matching function. Both claim the result of an undisclosed NLU system that receives a translated input, such as the text "I-want-frozen-peach-pie," and determines the nuances and intent of that request. No program of that kind, or any kind for that matter, is recited in the claims or the specification. So, in this respect, the fate of the "matching" and "identifying" claim terms are intertwined. And as with the matching function, the identify function—which identifies the specific product that the user wants—is a complex operation. The specification recites only the desired result: receive an utterance, translate it somehow, match it to some item somehow, and provide it to a user to purchase with a quote. This is not an invention that our patent laws protect. It is an aspiration, not a solution. It is an attempt to claim the idea of voice commerce and preempt an entire existing and yet-to-emerge means of commerce while providing no advance at all to the underlying technologies.

B. Several Terms in the Asserted Claims Are Ambiguous, and Therefore Indefinite Under Section 112 ¶ 2.

The Patent Act requires that a patent specification "conclude with one or more claims *particularly pointing out and distinctly claiming* the subject matter which the applicant regards as [the] invention." *Nautilus, Inc. v. Biosig Instruments, Inc.*, 572 U.S. 898, 901 (2014) (citing 35 U.S.C. § 112 ¶ 2). A claim is indefinite when, read in light of the specification and the prosecution history, it fails to inform, with reasonable certainty, those skilled in the art about the scope of the invention. *Id.* at 902; *see also Media Rights Techs., Inc.*, 800 F.3d at 1371 ("Notably, a claim is indefinite if its language 'might mean several different things and no informed and confident choice is available among the contending definitions.'") (quoting *Nautilus*, 572 U.S. at 911 n. 8). Several terms render the asserted claims in which they appear indefinite.

1. "placing more weight on words"

Amazon's Construction	Freshub's Construction
Indefinite under 35 U.S.C. § 112 ¶ 2	assigning weight to words based on user's preferences, the type of item in the spoken order, or the user's history

All the asserted patents contain asserted claims that recite "placing more weight on words" as part of at least one claimed function. (*See* '094 patent cl. 6; '153 patent cl. 9; '408 patent cls. 5, 6; '810 patent cls. 5, 6.) The patents are entirely ambiguous as to how word-weighting is expected to operate, when it is performed, or on which words "more weight" is placed. And to make matters worse, the phrase is used inconsistently across the claims. Some claims use "word-weighting" during translation (ASR):

> The voice recognition system as defined in claim 1, wherein the system is configured to increase recognition accuracy by placing more weight on words related to types of items being purchased by the user in *translating at least a portion of the digitized order to text*.

('810 patent cl. 5, emphasis added.) Some claims use "word-weighting" during the identification function (NLU):

> 6. The computer-implemented method as defined in claim 1, wherein identifying an item corresponding to the text further comprises placing more weight on words in the digitized order related to the user's past purchase history *when identifying the item corresponding to the text*" ('094 patent cl. 6)

Yet in other claims it is not at all clear whether the weighting is used in the ASR or NLU process or both:

> 9. The voice processing system as defined in claim 1, wherein the [computer] system is configured to increase recognition accuracy by, when there are more than one potential matches to the text, placing more weight on words related to the user's past purchase history." ('153 patent cl. 9; '408 patent, cl. 6; '810 patent cl. 6)
>
> 5. The voice processing system as defined in claim 1, wherein the system is configured to increase recognition accuracy by placing more weight on words related to types of items being purchased by the user. (''408 patent, cl. 5.)

The problem is that phrase "placing more weight on words" is meaningless without an algorithm that specifies what words are weighted and how those weights are used in the algorithm to either translate speech to text or to identifying items from that text. Like the terms discussed above, this phrase is nothing but an aspirational placeholder for "do it even better using past purchases or related items or whatever else might work." That too is not a solution. Nor is Freshub's proposed construction—"assigning weight to words based on user's preferences, the type of item in the spoken order, or the user's history—which merely swaps "placing" with "assigning," repeats the remainder of the term, broadens it by adding words, but fails to explain it. *See DeMarini Sports, Inc. v. Worth, Inc.*, 239 F.3d 1314, 1323–24 (Fed. Cir. 2001) ("The construction of claims is simply a way of elaborating the normally terse claim language in order to understand and explain, but not to change, the scope of the claims.") (internal quotation and citation omitted). How,

when, what words?

2. "the first system," "the first computer system," and "the second computer system"

Amazon's Construction	Freshub's Construction
Indefinite under 35 U.S.C. § 112 ¶ 2	Plain and ordinary meaning

Claim 20 of the '094 patent recites "a first computer" and "instructions that when executed by the first computer cause *the first system* to perform" certain operations, and "the remote system" "wherein *the second computer system* is configured to transmit the digital representation over the network to *the first computer system*." The terms "the first system," "the first computer system," and the "second computer system" each lack antecedent basis. The claim also raises several questions. Is the "first computer" the same or different from "the first system"? Is "the first system" part of "the first computer"? Does "the remote system" include "the second computer system" or are the two systems distinct? And how is "the first computer system" different from "the first computer" or "the first system"? Because none of these questions is answered by the claim, the claim is indefinite. *See Smith v. ORBCOMM, Inc.*, No. 2:14-cv-666, 2015 WL 5302815, *12–13 (E.D. Tex. Sept. 10, 2015) (holding claim invalid for indefiniteness based on lack of antecedent basis). And the claim cannot be corrected because it is subject to multiple interpretations and the terms are subject to a reasonable debate as to their meaning. *Id.* (citing *Novo Indus., L.P. v. Micro Molds Corp.*, 350 F.3d 1348, 1357 (Fed. Cir. 2003)). Indeed, the Federal Circuit "repeatedly and consistently has recognized that courts may not redraft claims, whether to make them operable or to sustain their validity." *Chef Am., Inc. v. Lamb-Weston, Inc.*, 358 F.3d. 1371, 1373–74 (Fed. Cir. 2004) ("Even 'a nonsensical result does not require the court to redraft the claims of [a] patent.'").

V. CONCLUSION

For the foregoing reasons, Amazon requests that the Court should find the disputed functional terms indefinite under § 112 ¶ 6 and the disputed terms that lack antecedent basis indefinite under § 112 ¶ 2.

Dated: January 24, 2020

Respectfully submitted,

/s/ J. David Hadden

J. David Hadden, CSB No. 176148
Email: dhadden@fenwick.com
Saina S. Shamilov, CSB No. 215636
Email: sshamilov@fenwick.com
Ravi R. Ranganath
Email: rranganath@fenwick.com
Vigen Salmastlian, CSB No. 276846
Email: vsalmastlian@fenwick.com
FENWICK & WEST LLP
Silicon Valley Center
801 California Street
Mountain View, CA 94041
Telephone: 650.988.8500
Facsimile: 650.938.520

Counsel for Defendants
AMAZON.COM, INC.,
AMAZON DIGITAL SERVICES, LLC,
PRIME NOW, LLC, and
WHOLE FOODS MARKET SERVICES, INC.

Of Counsel:

Barry K. Shelton (TX Bar #24055029)
bshelton@sheltoncoburn.com
SHELTON COBURN LLP
311 RR 620 S, Suite 205
Austin, TX 78734
Tel: (512) 263-2165
Fax: (512) 263-2166

CERTIFICATE OF SERVICE

The foregoing document was filed under the Court's CM/ECF system, automatically effecting service on counsel of record for all other parties who have appeared in this action on the date of such service.

/s/ J. David Hadden
J. David Hadden